基于条件生成对抗网络的梯级表面高光去除方法

胡宇航 [1]，胡海洋 [1]，李忠金 [1,2]

(1. 杭州电子科技大学 计算机学院，杭州 310018; 2. 浙江省北大信息技术高等研究院智能软件技术与应用研究中心，杭州 310000)

摘　要：针对传统高光去除方法对梯级图像高光去除表现不佳问题，提出了一种基于条件生成对抗网络的梯级图像高光去除方法(Multi-scale Spatial dense gradient cascade generative adversarial network，MSDGC-GAN)。该方法设计了一种空间上下文密集模块 (Spatial Contextual Feature Dense Block，SCFDB)能够深度提取像素行与列之间的空间背景信息。此外，设计了一种多尺度梯度级联结构以弥补网络下采样中的尺度特征损失，并且该结构能够赋予模型多尺度鉴别能力同时稳定训练梯度分布。在分析了经典双色反射模型基础上，将最大漫反射度估计应用于损失函数以监督网络训练。实验结果表明，所提方法在经典高光数据集和自制梯级高光图像数据集中表现均优于所对比方法。

关键词：高光去除；条件生成对抗网络；多尺度；特征级联；漫反射估计；密集连接网络

中图分类号：TP391.4　　doi: 10.19734/j.issn.1001-3695.2022.02.0089

Conditional generative adversarial network-based method for stepped surface highlight removal

Hu Yuhang[1], Hu Haiyang[1], Li Zhongjin[1, 2]

(1. *College of Computer & Technology*, *Hangzhou Dianzi University*, *Hangzhou* 310018, *China*; 2. *Intelligent Software Technology & Application Research Center*, *Advanced Institute of Information Technology*, *Peking University*, *Hangzhou* 310000, *China*)

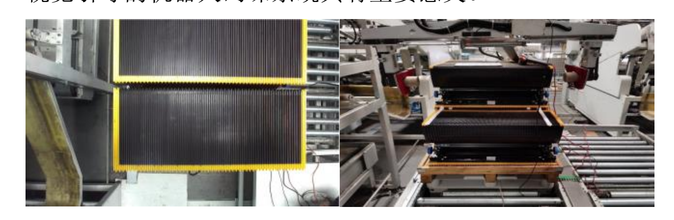
Abstract: It is difficult for traditional highlight removal algorithms to effectively deal with the processing of stepped highlight images in the stepped palletizing of factory robots. To solve this problem, based on the knowledge of conditional generative adversarial network, this paper proposes a stepped surface highlight removal network model named MSDGC-GAN (Multi-scale Spatial dense gradient cascade generative adversarial network) . In this method, the Spatial Contextual Feature Dense Block (SCFDB) aims to deeply extract the spatial background information between pixel rows and columns. In addition, the multi-scale gradient cascade structure aims to compensate for the scale feature loss in network downsampling, and this structure can endow the model with multi-scale discriminative ability while stabilizing the training gradient distribution. Based on the analysis of the classical two-color reflectance model, this paper apply the maximum diffuse reflectance estimation to the loss function to supervise the network training. The experimental results show that the proposed method outperforms the compared methods in both the classical highlight dataset and the self-made stepped highlight image dataset.

Key words: highlight removal; conditional generative adversarial network; multi-scale; feature cascade; diffuse reflectance estimation; densely connected network

0　引言

随着新一轮科技革命和产业变革浪潮的兴起，越来越多的企业开始注意到企业生产智能化一体化在未来竞争中所具备的独特优势，而机器人作为智能装备的代表，已越来越成为助推企业转型升级的动力器，企业生产效益的倍增器，企业竞争优势的放大器，并成为各大企业争夺经济发展的制高点，被广泛的应用在各大生产场景下。近年来机器视觉技术不断进步与创新，其在实际生产环境下的技术应用越来越成熟，以视觉引导为基础的机器人已经被广泛应用于各类自动化生产场景。利用机器视觉技术对抓取物体进行定位，再将相关的位姿信息传递给机器手使其具备“感知”能力，成为当前智能化工厂发展导向，而其中对于抓取物体图像的相关处理是机器人能否成功抓取的关键前提。

与普通材质物体不同，在实际扶梯梯级码垛生产线上，机器人所要抓取的梯级材质通常为金属铝制，为此在自然光照射下具有高反光、背景复杂等特性，如图 1(a)所示。梯级表面部分区域存在高光覆盖会给后续视觉引导中的图像处理步骤如阈值分割、边缘直线检测等带来麻烦，容易造成分割不匀、检测失败等问题，进而影响后续梯级识别、抓取点定位提取等。因此，金属材质的扶梯梯级图像高光去除对于以视觉引导的机器人码垛系统具有重要意义。



(a) 梯级表面高光　　(b) 机器人码垛工位

图 1　工厂梯级码垛环境

Fig. 1　Factory step palletizing environment

在早期，许多方法[1,2]采取多视图方式给图像增加额外的约束来实现高光去除，对于硬件操作要求较高，而其他方法则基于单视图对图像进行高光去除。然而此类算法大多基于

收稿日期：2022-02-25；**修回日期：**2022-05-03

作者简介：胡宇航(1995-)，男，浙江慈溪人，硕士研究生，主要研究方向为工作流识别技术、图像算法(192050155@hdu.edu.cn)；胡海洋(1977-)，男，江苏宝应人，教授，博士，主要研究方向为软件业务过程技术、过程检索与挖掘；李忠金(1986-)，男，江苏句容人，讲师，博士，主要研究方向为云计算、工作流调度.

严格的假设分析和先验条件约束，如颜色空间分析和稀疏矩阵分解，在图像纹理或者光照条件复杂的情况下算法性能差，不能满足实际需求。近年来，随着深度学习发展，其已广泛加入社会生产、生活的各个方面。与传统算法相比，深度学习能够通过对训练图像的自学习捕捉图像的相关特征信息，使得其在图像领域也迈上了一个新的台阶，不仅在高级语义视觉任务中如目标检测、文字识别等取得了成功[3]，其在各种低级视觉任务中也取得了成功。然而，由于缺乏大量高质量的高光训练数据，目前利用相关深度学习方法进行高光去除的研究仍然较少，通常仅局限于自制小规模数据集进行训练，不具备通用解决性，如 Yi 等人[4]将对象外观分解为高光、阴影和反照率层，通过在线收集产品照片来构建多视图数据集对图像进行高光去除。高光去除方法可以看成是一种图像复原生成方法，而近年来生成对抗网络(generative adversarial network)的提出与发展，使得其在图像复原生成领域如图像去雾[5,6]、去噪声[7]和阴影去除[8,9]有着突出表现，效果要优于传统卷积网络模型。因此生成对抗网络的发展也为图像高光去除提供了一个全新的研究方向。

综上所述，根据实际工厂需要，为提高梯级图像获取质量从而更利于后续视觉引导系统的分析与处理，本文提出一种基于条件生成对抗网络的梯级图像高光去除模型 MSDGC-GAN(Multi-scale Spatial dense gradient cascade generative adversarial network)，如图 2 所示。其目的在于尽可能去除梯级表面高光信息，同时最大程度还原梯级纹理细节，提高整体图像质量。首先，采用编码-解码(encode-decode)结构作为生成器基本架构，在密集连接卷积网络基础上(densely connected convolution network)构造了一种 SCFDB 模块(Spatial Contextual Feature Dense Block，空间上下文密集块)作为生成器的基本模块，模块中基于多路并联片间(slice-by-slice)卷积来提取和传递图像像素背景特征信息，通过深层次的密集网络获取图像的语义特征信息；其次，提出一种自底向上的特征级联方式，旨在弥补网络下采样时所损失的部分尺度信息。在借鉴 MSG-GAN 网络[10]基础上，提出一种多尺度梯度鉴别方法，将编码器各尺度级联后的特征图输出通过 SOS(Strengthen-Operate-Subtract)增强策略结构与解码器长距离相连，其分支经卷积后输出至对应鉴别器；最后，在损失函数构造上分析了传统双色反射模型，将最大漫反射分量估计应用于损失函数的监督训练。由于目前无大量梯级表面高光图像数据，为了能够训练本文网络框架并评估它的优势，本文创建了一个一对一的梯级表面高光对照数据集，数据集分为高光组和无高光组。实验表明，本文提出的梯级表面高光去除网络框架 MSDGC-GAN 的表现要明显优于传统高光去除算法。

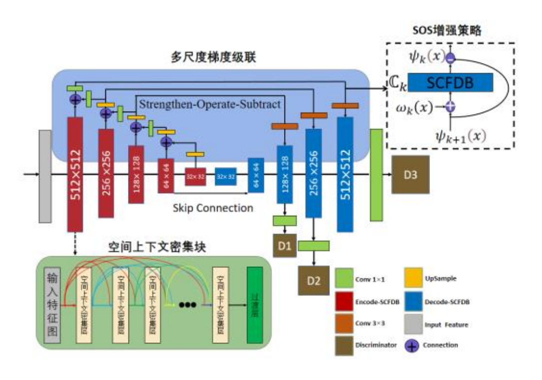


图 2 MSDGC-GAN 高光去除模型

Fig. 2 MSDGC-GAN highlight removal model

本文的其余部分组织如下：第一节对现阶段图像高光去除算法和生成对抗网络进行相关介绍；第二节对本文所提方法从网络架构、模块组成和损失函数三方面进行阐述；第三节利用自定义数据集对本文方法和传统方法进行结果对比并实现消融实验；第四节对全文进行总结，并对未来的研究方向进行了展望。

1 相关工作

1.1 图像去高光算法

对于图像去高光而言，现阶段大多数的处理方式可分为单视图和多视图两种方式。前者认为物体表面的光照可由漫反射分量和镜面反射分量所构成，其主要思路也集中在如何最大程度地分离图像的镜面反射分量。Tan 等人[11]开创性提出先估计伪无镜图像，然后通过比较生成的伪无镜图像与输入图像的强度对数微分，迭代去除镜面分量。Yang 等人[12,13]对该方法进行了扩展，提出用双边滤波方法对伪无镜面图像和输入图像进行比较，达到图像实时处理目的。由于这些伪无镜图像会显著影响最终的光点去除结果，因此基于先验假设理论的高光去除方法也应运而生，如 Ramos 等人[14]提出了一种从单个图像中分离漫反射和镜面反射分量的全自动方法以更少的计算成本获得了更好的质量结果。Yamamoto [15]等人对现有单张图片反射分量分离方法进行改进进一步提高检测精度。Ye[16]等人提出一种基于暗通道先验的镜面高光图像增强算法，通过引入基于局部像素色差的加权函数来解决图像中的光晕伪影。郭碧茹等人[17]提出了一种导向滤波的高光去除改进算法，通过设定阈值分离图像黑色像素同时通过导向滤波的优势最大限度保留图像的边缘和纹理避免了纹理效应。但是，由于这些方法需要严格的先验假设，在背景和光照条件复杂的真实场景中，这些方法很容易失败，导致高光无法有效去除问题。

而基于多视图的去高光方式认为高光是方向相关的，所以依赖于从多视角多副图像寻找图像特征关系并进行相互映射以施加额外的约束从而达到图像去高光复原的效果。如文献[1,2]选取固定的场景从不同的角度拍摄了多副图像，而 Li 等人[18]则在不同光源下对物体拍摄一系列图像，Shah 等人[19]提出利用视频片段中相邻图像帧中的附加信息来减少每个帧的镜面反射从而实现高光去除。温佩芝等人[20]提出了一种基于多视角图像特征匹配的高光去除方法。此类多视图方法虽然能达到不错的去除效果，但是其缺陷也明显，即在实际过程中需要有条件的一系列图像作为支撑，应用拓展性不好。

1.2 生成对抗网络

生成对抗网络由 Goodfellow 等人[21]提出，主要思想来源于二人零和博弈，其网络结构由生成器与鉴别器所构成。生成器接受一个随机噪声生产对应的假样本，而鉴别器则接受生成器所给的假样本和真样本并试图判别他们的真伪。总的来说生成网络是为了生成更具欺骗性的样本使得鉴别器无法分辨真伪，从而达到以假乱真的效果，使得整体处在一个动态平衡过程，其理论关系可表示为

$$\min_G \max_D V(D,G) = \mathop{E}_{x\sim P_r}[\log(D(x))] + \mathop{E}_{z\sim P_g}[\log(1-D(G(z)))] \tag{1}$$

其中，P_r 代表真实数据的样本分布，P_g 则代表生成器产生的数据分布，z 代表简单样本噪声。而由于传统生成网络无法控制图像的生成，M Mirza 等人[22]提出了一种条件生成对抗网络 CGAN(Conditional Generative Adversarial Nets)，其将原生 GAN 网络的概率判断全改成条件概率，即在生成器与判别器的输入端增加条件输入以指导图像生成，损失函数表示为

$$\min_G \max_D V(D,G) = \mathop{E}_{x\sim P_r}[\log(D(x|y))] + \mathop{E}_{z\sim P_g}[\log(1-D(G(z|y)))] \tag{2}$$

其中，y 表示先决条件。基于 CGAN 思想，Isola 等人[23]提出了一种有监督的图像到图像翻译网络(pix2pix)，用于图像配对的翻译转换，其生成器与鉴别器分别使用了 U-net 架构和马尔可夫判别器。pix2pix 方法取得了良好的效果，但是由于

其网络结构问题最终生成图片只适合低分辨率图像，并且与真实图像存在一定差距，为此 Wang 等人[24]提出一种高分辨率 pix2pixHD 网络，利用多尺度判别技术弥补了 pix2pix 网络对于生成高分辨率图像的不足，但其生成的图像色彩过大，虽然对于粗尺度上的图像生成能够有效的进行监督，但对于细尺度的生成具有缺陷难以生成具有质感的真实图像。Zhu 等人[25]提出一种无监督的循环对抗网络(CycleGAN)，解决了图像到图像转换需要配对图像的问题。为了解决 GAN 网络训练时梯度不稳定和对超参数敏感的问题，A Karnewar 等人[10]提出了一种多尺度梯度生成对抗网络(MSG-GAN)，将鉴别器与各尺度特征输出图建立通道从而稳定梯度流动。

为了解决传统高光去除算法对于梯级表面高光处理效果不佳的问题，本文基于条件生成对抗网络提出了一种 MSDGC-GAN 梯级高光去除模型，利用自定义梯级数据集训练模型网络，在实验分析中本文不仅与传统高光去除算法进行对比，同时加入了基于 pix2pixHD 的模型方法进行比较，实验证明本文所提出 MSDGC-GAN 模型表现要优于现有方法。

2　本文方法

2.1　网络架构

本文的目的是通过输入一张待去除高光梯级图像 I_H，且不通过其他任何相关信息辅助生成一张高光抑制的对应清晰图像 I_S，因此构建了一个多尺度条件生成对抗网络，旨在通过端到端的方式获取高光去除图像。网络主体结构由生成器与鉴别器组成，其中生成器基本架构采用 U-net 结构，利用其编码-解码结构特点提取图像深层结构信息。同时为了充分提取特征像素之间的空间上下文背景信息特征，提出了一种空间上下文密集块作为生成器的基本模块。为了解决网络在下采样池化操作中容易丢失部分尺度特征信息的问题，提出一种多尺度梯度级联方法。通过从底层特征依次进行级联输出以弥补相邻模块之间的下采样特征损失，并将各尺度鉴别器与生成器级联输出各自相连，使得网络增强对图像细节的处理能力并具备多尺度鉴别能力，实现了梯度的传导改善了网络训练时梯度不稳定的问题。在损失函数阶段，分析了双色反射模型，将图像的漫反射分量估计应用于损失函数，同时结合对抗性损失函数和特征匹配损失作为目标总损失。网络的总体架构如图 2 所示。

2.2　生成器架构

2.2.1 空间上下文密集块

本文生成器结构以 U-net 网络为基础，同时为了解决传统 U-Net 网络在网络层数加深情况下梯度消失和表面特征易丢失的问题，本文借鉴 DensNet[26]网络理念，将 DB 密集块(Dense Block)引入 U-net 网络，通过密集连接来加强特征和梯度的传递，如图 3 所示。但是传统密集块的堆叠卷积方式不能有效的获取各行各列像素间的空间背景信息，而这种信息已经被证明对高亮检测和低级图像处理是有用的[27]。

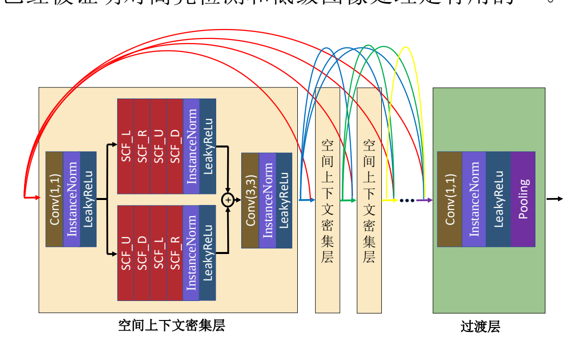


图 3　空间上下文密集块

Fig. 3　Spatial context dense block

$$S_i' = \begin{cases} R(G(S_i)) & i=1 \\ S_i' + R(G(S_i)) & i=2,3,\ldots,H \end{cases} \tag{3}$$

其中，G 为卷积操作，S_0 为卷积后得到的特征切片，R 为激活函数采用 LeakyReLu。其中每个维度方向上的片间卷积函数共享同一个卷积核，通过这种方式可以使得更加丰富的特征信息在像素维度的层与列中实现传递从而有效提取像素的空间背景信息，并且其效率优于普通卷积，简化了信息传递并加快模型的运算效率。对每一路切片信息进行拼接和 LeakyReLu 激活后将两路支路进行加权融合，最后经过一次 3×3 卷积层输出特征图，如图 4 所示。

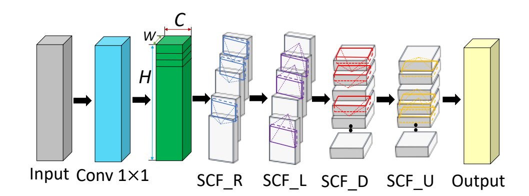


图 4　Slice-by-Slice 卷积示意图

Fig. 4　Slice-by-Slice convolution diagram

本文密集块之内用 SCF 层对相同尺寸的输入特征图进行特征提取。与传统密集层类似，每一个 SCF 层的输出都会被添加到后续所有 SCF 层作为共同输入，同时编码器密集块之间利用过渡层的 1×1 卷积压缩特征图通道，再通过平均池化实现特征图的下采样将源输入特征图每次降至 1/2 尺寸。解码器上采样过程采取与编码器相同的密集层连接结构，而在过渡层中则使用 4×4 转置卷积(Transpose Convolution)将特征图依次向上采样。层内所有归一化采取实例化归一化(InstanceNorm)技术。

2.2.2 多尺度梯度级联

如图 2 所示，本文将 SCFDB 模块作为生成器的基本模块，整个编码器的下采样过程涵盖 5 个密集块，每个块内有 4 个密集连接层。虽然通过密集块能够在每个层级极大丰富特征信息，但是无法解决 U-net 网络在下采样操作过程中存在容易丢失部分尺度特征信息的问题，且网络的底层特征已被证明在恢复物体细节方面具有重要作用。为此本文提出一种特征级联方法，通过从底层特征逐步与上层特征在通道维度进行级联以增强尺度语义特征的传递，如图 5 所示。

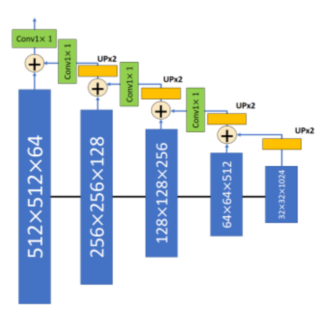


图 5　特征级联方法

Fig. 5　Feature cascading method

通常网络的深层特征图具有较高的通道数，如直接进行特征图的堆叠拼接将会增加网络运行参数影响网络运行效率。为了减少网络内存的消耗，本文通过一次简单卷积操作将每次级联后的特征块通道数压缩至原始数量级。设在编码器下采样操作后得到第 k 个尺度的特征块为 $\omega_k(x)$，则每一级级联操作可表示为

$$\begin{aligned} \tilde{\omega}_k(x) &= Conv(Up(\omega_k(x))) \\ \tilde{\omega}_i(x) &= Conv(Up(Cat(\omega_i(x), \tilde{\omega}_{i+1}(x)))) \quad i=1,\ldots k-1 \end{aligned} \tag{4}$$

其中，$\tilde{\omega}_k(x)$ 为级联后输出特征块，$Up(\ )$、$Cat(\ )$ 和 $Conv(\ )$ 表示上采样、串联和卷积操作。同时受 MSG-Net 启发，本文在特征级联方法上提出了一种多尺度梯度级联方法(图 2)，将编

码器每个尺度级联后的特征输出图经过一次 3×3 卷积，随后通过长距离跳跃连接送入对应尺度的解码器转置后特征块。此外解码器在不同尺度的层级上有分支，将每次尺度级联增强后的特征图进行输出，经过一次 1×1 卷积后将特征图像送入对应尺度鉴别器，使得鉴别器不但以最终输出作为输入，同时还以编码器中间级联输出作为输入。这让模型能够在更多尺度上捕捉图像信息，使得最终生成图像具有更好的表征同时让网络梯度可以直接从鉴别器流向生成器的中间层，增加了训练期间的稳定性，并解决了 U-Net 架构非常深入时梯度渐变消失的问题。本文不采用传统 U-net 的长距离跳跃连接方式将模块中的上采样增强特征和潜在特征连接起来，而是采取 SOS 增强策略结构[28]，此结构受图像增强算法启发设计能够对图像进行细化增强处理。如图 2 虚框所示，其操作可表述为

$$\psi_k(x)=\mathbb{C}_k(\omega_k(x)+Up(\psi_{k+1}(x)))-Up(\psi_{k+1}(x)) \tag{5}$$

其中，$\psi_k(x)$ 表示解码器输出特征图，$\mathbb{C}$ 表示修复单元，每个修复单元代表一个密集块。对于第 k 层的输出 $\psi_k(x)$，首先对上一层得到的特征图 $\psi_{k+1}(x)$ 进行 2 倍的上采样，然后将对应编码器输出的特征图 $\omega_k(x)$ 相加送入修复单元，其结果再减去 $\psi_{k+1}(x)$ 作为最终输出，通过这种方式细化输出结果以增强图像细节。

2.3　鉴别器架构

在生成器训练过程中解码器将会输出各个尺度的图像，本文采用 PatchGAN[23]判别器结构作为鉴别器网络。相比于传统二分类的鉴别器，PatchGAN 将图片分为多个区域分别进行单独判别并输出预测概率值，而不是直接对整幅图片进行打分，从而在图像局部细节特征的提取上具有更好的表达能力。本文将源图像缩放至与编码器分支输出图像相应尺度，并互相连接起来，作为一个假的样本提供给鉴别器，同样另外将源图像和实际目标图像同尺度也相互连接作为真样本提供给鉴别器。本文使用了 3 个鉴别器 D1, D2, D3 来接受不同尺度的输出图像，所有鉴别器的结构都是相同的。

2.4　损失函数

本文网络训练的目的是尽可能将输入图片转换为目标无高光图，为此提出一种漫反射分量损失函数(Diffuse reflection loss)，即将漫反射分量估计应用于内容损失用来约束预测图与目标图的差异。根据文献 11 提出的双色反射模型，非均匀光照下的任意一点的光照反射可分为漫反射与镜面反射，其数学模型为

$$P(Z)=w_f(Z)F(Z)+w_h(Z)H(Z) \tag{6}$$

其中，$P(Z)$ 是图像像素强度，$w_f(Z)$ 为漫反射分量系数，$w_h(Z)$ 为镜面反射分量，$F(Z)$ 为漫反射分量，$H(Z)$ 为镜面反射分量。其中图像中任意一点像素点的色度可表示为

$$\begin{cases}\alpha(Z)=\dfrac{P(Z)}{P_r(Z)+P_g(Z)+P_b(Z)}\\ \beta(Z)=\dfrac{F(Z)}{F_r(Z)+F_g(Z)+F_b(Z)}\\ \gamma(Z)=\dfrac{H(Z)}{H_r(Z)+H_g(Z)+H_b(Z)}\end{cases} \tag{7}$$

$$w_f(Z)F(Z)=P(Z)-\frac{\beta_{max}(Z)(P_r(Z)+P_g(Z)+P_b(Z))-P_{max}(Z)}{3\beta_{max}(Z)-1} \tag{8}$$

因此，若能求解最大漫反射色度 $\beta_{max}(Z)$，便能获得图像的漫反射分量，根据文献[13]所提通常最大漫反射分量可由式(9)表示：

$$\tilde{\beta}(Z)=\frac{\alpha_{max}(Z)-\alpha_{min}(Z)}{1-3\alpha_{min}(Z)} \tag{9}$$

其中，$\tilde{\beta}(Z)$ 是最大漫反射色度 $\beta_{max}(Z)$ 的估计值,因此定义多尺度下的漫反射分量损失函数如下：

$$\mathcal{L}_D=\frac{1}{2n}\sum_{k=1}^{n}\frac{1}{c_k w_k h_k}\|z_k-y_k\|^2 \tag{10}$$

其中，z_k 和 y_k 分别为第 k 尺度图像的预测和真实的漫反射分量，c_k，w_k 和 h_k 为第 k 尺度图像尺寸大小，漫反射分量可由式(8)(9)求出。

本文总的结构损失函数由对抗性损失函数、漫反射分量损失函数和特征匹配损失三个部分所组成，其中采用 PIX2PIX-HD 所提出的改进对抗损失，其公式如下：

$$\mathcal{L}=min_G\left(max_{D_1,D_2,..D_n}\sum_{k=1}^{n}\mathcal{L}_{GAN}(G,D_k)\right) \tag{11}$$

其中 $\mathcal{L}_{total}=min_G\left(max_{D_1,D_2,..D_n}\sum_{k=1}^{n}\mathcal{L}_{GAN}(G,D_k)\right)+\lambda_1\sum_{k=1}^{n}\sum_{i=1}^{T}\frac{1}{N_i}\left[\left\|D_k^{(i)}(x_k,y_k)-D_k^{(i)}(x_k,z_k)_1\right\|_1\right]+\lambda_2\frac{1}{2n}\sum_{k=1}^{n}\frac{1}{c_k w_k h_k}\left\|D_k^{(i)}(x_k,y_k)-D_k^{(i)}(x_k,z_k)_1\right\|_1$ 表示条件损失函数：

$$\mathcal{L}_{GAN}(G,D)=E_{(x,y)}[\log D(x,y)]+E_{(x,z)}[\log(1-D(x,G(x,z)))] \tag{12}$$

关于特征匹配损失部分，其定义如下：

$$\mathcal{L}_{FM}(G,D_k)=\sum_{k=1}^{n}\sum_{i=1}^{T}\frac{1}{N_i}\left[\left\|D_k^{(i)}(x_k,y_k)-D_k^{(i)}(x_k,z_k)_1\right\|_1\right] \tag{13}$$

其中，N_i 代表判别器第 i 层的元素数，T 为总层数，$D_k^{(i)}$ 代表判别器 D_k 的 i 层特征。

总的损失函数如下：

$$\begin{aligned}\mathcal{L}_{total}=&min_G\left(max_{D_1,D_2,..D_n}\sum_{k=1}^{n}\mathcal{L}_{GAN}(G,D_k)\right)+\\&\lambda_1\sum_{k=1}^{n}\sum_{i=1}^{T}\frac{1}{N_i}\left[\left\|D_k^{(i)}(x_k,y_k)-D_k^{(i)}(x_k,z_k)_1\right\|_1\right]+\\&\lambda_2\frac{1}{2n}\sum_{k=1}^{n}\frac{1}{c_k w_k h_k}\|z_k-y_k\|^2\end{aligned} \tag{14}$$

其中，λ_1、λ_2 分别为特征匹配损失函数和漫反射损失的权值。

3　实验结果与分析

3.1　实验设置

本文去高光条件生成对抗网络基于 Pytorch 深度学习框架搭建，编程语言为 Python3.7，网络训练服务器配置为八核 Inter CPU I7，图像处理器(GPU)采用 NVIDIA GTX 2080Ti，显存 20GB。在训练时采用自适应动量估计优化算法(Adam)作为求解器，动量参数 β_1 为 0.5，β_2 为默认值，权重采用高斯分布随机初始化，均值为 0，标准差为 0.02，总共训练 200 个 epoch，前 170 个 epoch 保持初始学习率不变，后 30 个 epoch 采取线性衰减方式至 0。对于损失权重的权值，经过多次实验设置 λ_1 为 10，λ_2 为 0.5。由于大多数图像高光去除的传统算法基于颜色空间分布和矩阵运算原理，这类算法不需要大量的图片进行验证，为此目前还未有大规模的公共数据库用于高光梯级图像的去除，因此本文对现实梯级进行拍照采集建立了用于训练与测试的数据集。为了模拟高光效果，通过打光器对梯级进行照射分别收集了在同一位置高光照射下和与无高光照射的梯级物体图像，并对图像进行相同切割选出最优图像对，将其统一裁剪为 512×512 大小后(图 6)按照有无高光进行分组，总计 2000 组对照图像。

由于本文模型采用了全卷积结构因此对任何图片输入都适用，为了增加网络的泛化性和通用性，本文还从文献[11~16,29]中收集高光图像，并将其进行了数据集扩充处理操作至共 700 组用于泛化训练并且对其进行了分析比较。本文采取客观和主观方面对图像进行实验分析，客观评价方面选取了峰值信噪比(PSNR)和结构相似性(SSIM)作为分析评价指标，PSNR 越大表示失真越小，SSIM 越大表示图片越接近原始图像，同时本文针对所提模块和损失函数做了消融实验分析。

3.2　结果对比分析

为了评估本文所提出的 MSDGC-GAN 方法去高光效果，

首先选取了测试梯级高光图像进行了实验，并与文献[12]、文献[14~16]、文献[30]和基于 pix2pixHD 的模型方法分别做了对比，实验效果如图 7 所示。

可以看出传统基于颜色分析和优化的算法在实际处理梯级单一表面大面积高光时效果很差，如图 7 所示，文献[14]方法对于此类图像的高光去除方法较差，异常像素面积大，高光区域并未还原，说明算法存在缺陷。文献[16]方法在一定程度上对光点附近进行了抑制，但无法去除中心高光。而文献[30]方法则无法很好检测出此类高光，处理结果与输入图并无多大差距。Pix2pixHD 方法结果图颜色失真严重，高光去除效果差。本文模型在梯级的纹理细节处理上更为细腻，一定程度上保留了梯级纹理，同时颜色保真好，高光去除结果自然且未出现异常像素问题。表 1 是在本文梯级数据集上平均 PSNR 和 SSMI 的指标对比，可以本文模型在梯级测试集上的平均 PSNR 领先其余方法近 11/dB，而传统算法指标表现普遍较差，进一步显示了本文模型在处理梯级高光图像上的优势。

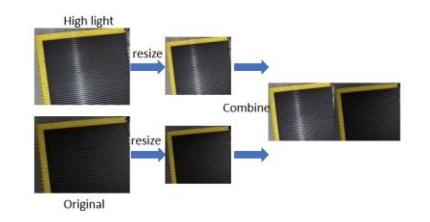
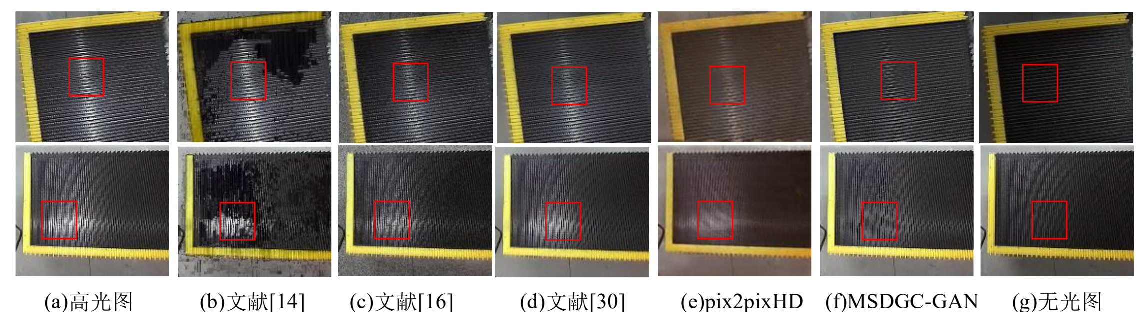


图 6　梯级数据集制作

Fig. 6　Cascade dataset production



(a)高光图　(b)文献[14]　(c)文献[16]　(d)文献[30]　(e)pix2pixHD　(f)MSDGC-GAN　(g)无光图

图 7　梯级表面高光去除对比

Fig. 7　Step surface highlight removal comparison

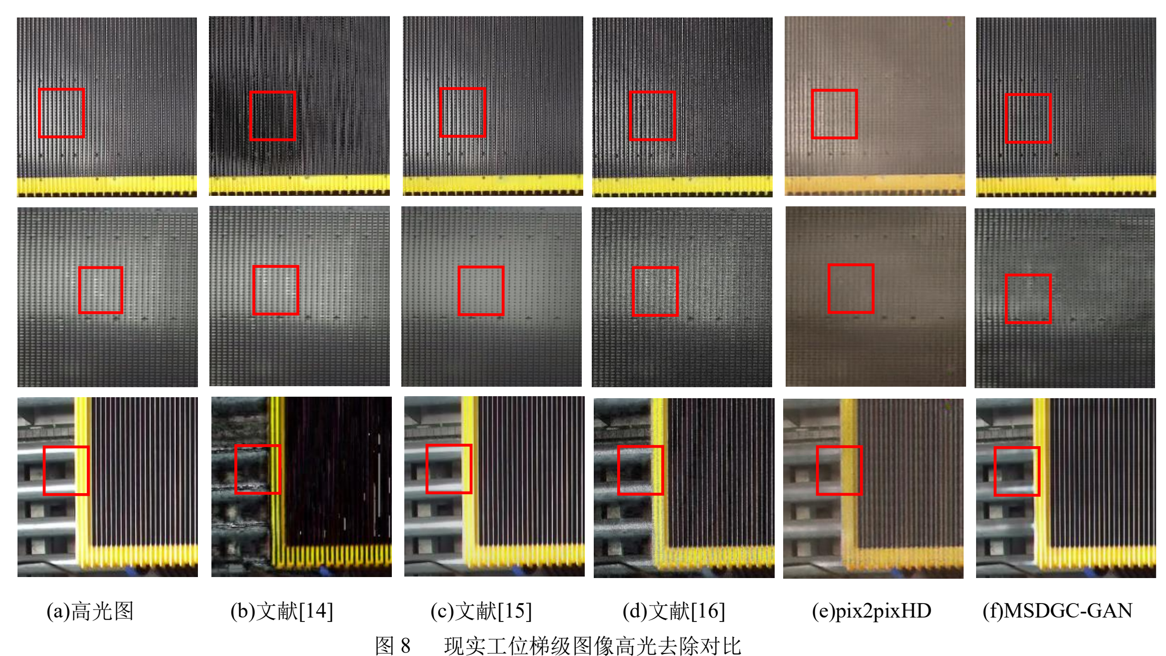
表 1　在梯级数据集中与不同高光去除方法的比较

Tab. 1　Comparison with Different Specular Removal Methods in the Cascade Dataset

方法	PSNR/dB	SSIM
文献[14]	18.281	0.672
文献[16]	20.512	0.772
文献[30]	22.675	0.753
Pix2pixHD	21.218	0.646
MSDGC-GAN	**29.298**	**0.842**

注：加粗字体为每列最优的结果。

图 8 为不同高光去除方法与本文方法在现实工位上收集的梯级图像结果对比图。文献[14]方法在第一行图片中对高光区域进行了处理，但颜色还原不自然，文献[16]方法对于高光的去除效果不明显。同时可以看出在处理背景金属高光区域时(第三行)，传统算法显得无能为力，既无法对梯级边缘高光进行抑制去除，也无法有效对背景强光进行消除，同时在背景处产生了大量噪点。而基于 pix2pixHD 的方法虽然对高光进行一定程度上抑制但是色彩过大不真实。由于高强度反光区域已经完全丧失了其原有应该具有的特征，算法只能通过附近像素信息尽可能来进行后续的消除还原，而本文模型通过 SCFDB 模块的稠密特征提取，不仅能够对梯级黄色表面高光进行抑制，同时对于背景处的高强度金属高光消除也具有良好的呈现效果，说明其能有效提取和利用像素背景间的信息，图像整体结构保留完善，未发生严重颜色畸变，还原效果较好，足够满足实际生产中的后续处理要求。



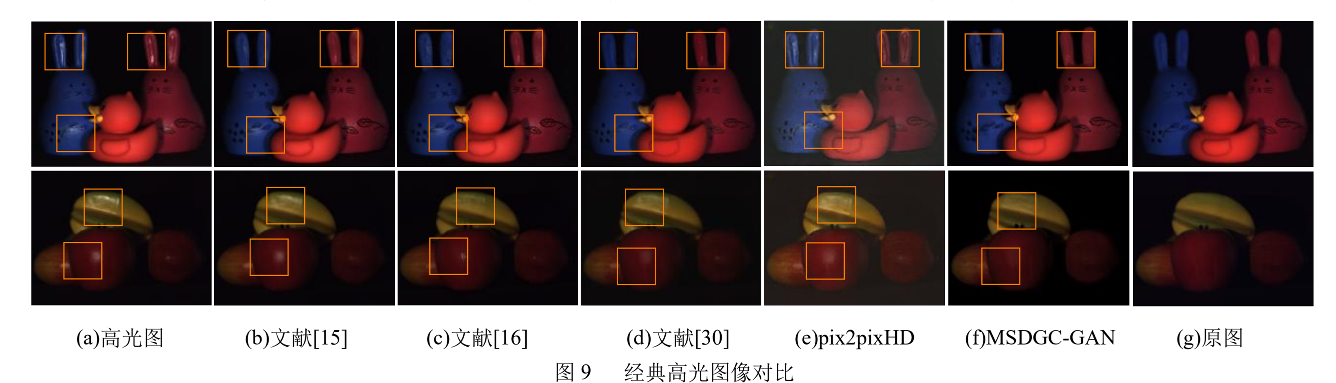
(a)高光图　(b)文献[14]　(c)文献[15]　(d)文献[16]　(e)pix2pixHD　(f)MSDGC-GAN

图 8　现实工位梯级图像高光去除对比

Fig. 8　Comparison of highlight removal in real work station step images

为了验证网络的通用泛化性，本文在训练中将经典高光图像进行数据集扩充并加入训练，选取了几张代表性高光图

像进行了对比实验，如图 9 所示。从视觉上分析各方法差距很小，其中在图 9(a)中，文献[15]方法在兔子耳朵和黄水果上方的光晕残留较，文献[30]方法水果高光去除结果中苹果存在光晕残留且还原较为模糊。表 2 为各方法在经典高光数据集上的平均 PSNR 和 SSIM 数值比较，可以看出 pix2pixHD 方法的平均性能最差，本文方法性能最优，说明模型在此类简单问题图像上表现也十分优异，整体视觉展示效果和色彩度还原更好，生成图像质量高。



(a)高光图　(b)文献[15]　(c)文献[16]　(d)文献[30]　(e)pix2pixHD　(f)MSDGC-GAN　(g)原图

图 9　经典高光图像对比

Fig. 9　Classic highlight image comparison

表 2　在经典高光数据集中与不同高光去除方法的比较

Tab. 2　Comparison with Different Specular Removal Methods in Classic Specular Datasets

方法	PSNR/dB	SSIM
文献[15]	38.342	0.912
文献[16]	39.876	0.932
文献[30]	40.108	0.938
Pix2pixHD	35.586	0.815
MSDGC-GAN	**40.298**	**0.942**

注：加粗字体为每列最优的结果。

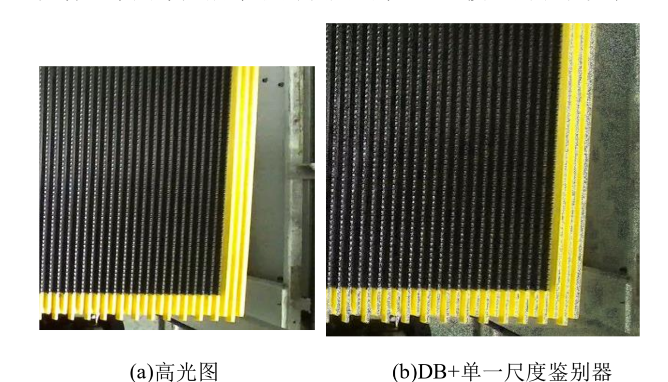
综上所述，本文所提方法不论是视觉效果还是指标比较上都有巨大优势，这说明了本文网络所采用的深度编码-解码结构不仅有效利用了 U 型对称网络结构提取深层次信息的优势，同时通过空间上下文密集模块与多尺度梯度级联增强了视野范围，充分提取了图像的纹理、背景特征信息，为此能够有效生成较高质量的图像，使表面特征得到有效恢复，特征还原更接近实际图片。

3.3　消融实验

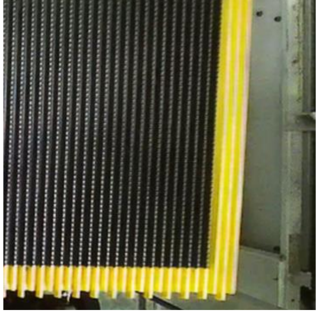
为验证 SCFDB 模块和多尺度梯度级联对于图像结果的影响，本文将网络按照不同模块组合在梯级图像集上进行消融对比，共设计为三组：第一组为 DB 模块作为模型基本模块，采用单一尺度鉴别器方式，第二组为采用 DB 模块和本文多尺度梯度级联方式，第三组为 SCFDB 模块和多尺度梯度级联，实验结果如图 10 所示。可以看出只采用传统密集块时，如图 b 所示，由于没有利用图像多尺度信息，结果图较为模糊，梯级纹理还原度不够好质量较差。而对于图 10(c)，图像对梯级边缘处还原不够细腻，出现了较为明显的像素异常现象，图像还原差。在应用本文模块后，不仅对于高光区域得到了有效抑制，上述问题也得到解决，图像更为逼真，证明了本文模块能更好捕获和传递图像像素间的特征并且能有效利用多尺度信息。

表 3 是本文模块消融实验在梯级数据集上的实验对比，通过实验结果可以看出，由于第一组方法并未采用多尺度鉴别器，模型运行速度相对最快，但指标表现最低。比较第 2、3 组实验可以发现，在同样采用多尺度梯度级联方法下，本文所提出的空间上下文密集块网络模型运行速度要快于传统密集块网络模型，这是由于 slice-by-slice 的卷积方式极大提高了卷积效率，且在 PSNR 和 SSMI 指标表现上都优于其他实验，进一步证明了本文方法的有效性。

图 11 显示了本文所提特征损失函数与漫反射损失函数在训练中的收敛情况，可以看到随着网络训练次数的增加模型逐渐趋于收敛，证明损失函数对生成器的训练起到了监督作用。同时为了研究损失函数对于图像重建效果的影响，基于梯级数据集，选取了不同的损失函数进行了消融实验：分别将漫反射损失替代为感知损失 $\mathcal{L}_P$，像素距离损失 $\mathcal{L}_1$ 并形成不同组合方式对其图像结果进行了指标对比，其中对抗损失和特征函数作为基准不单独进行消融实验。结果如表 4 所示，可以看出采用本文所采取的损失函数组合，模型的表现更佳。



(a)高光图　(b)DB+单一尺度鉴别器



c)DB+多尺度梯度级联　(d)SCFDB+多尺度梯度级联

图 10　不同模块组合效果

Fig. 10　Combination effects of different modules

表 3　在梯级数据集中不同模块组合比较

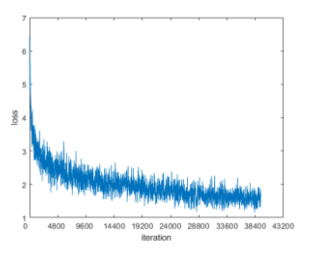
Tab. 3　Comparison of different module combinations in the rung dataset

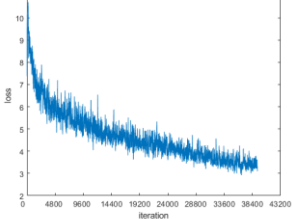
方法	PSNR/dB	SSIM	Time/s
DB+单一尺度鉴别器	27.341	0.826	**0.33**
DB+多尺度梯度级联	28.151	0.753	0.55
SCFDB+多尺度梯度级联	**29.751**	**0.853**	0.42

注：加粗字体为每列最优的结果。

为验证编码器与鉴别器多尺度梯度相连的影响对于训练时梯度分布的影响，本文分别采用编码器中间层级联输出作为鉴别器输入和不采用中间层级联输出分别训练 20 epoch 时的梯度分布，如图 12 所示，可以注意到，在不使用中间层输出时，由于梯度递减问题，梯度非常接近于零，而在使用编

码器中间层级联输出时，模型的梯度分布更好，原因是鉴别器与生成器相连实现了梯度的传导，缓解了梯度训练不稳定问题。





(a)特征损失函数　　(b)漫反射损失函数

图 11　梯级数据集损失函数收敛情况

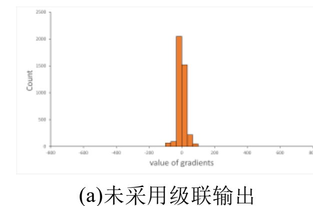
Fig. 11　Cascade dataset loss function convergence

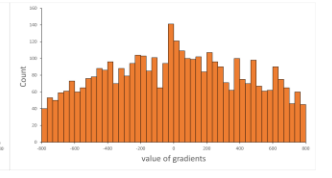
表 4　不同损失函数组合比较

Tab. 4　Comparison of different loss function combinations

方法	PSNR	SSIM
$\mathcal{L}_{GAN}+\mathcal{L}_{FM}$	27.123	0.818
$\mathcal{L}_{GAN}+\mathcal{L}_{FM}+\mathcal{L}_{D}$	**28.751**	**0.853**
$\mathcal{L}_{GAN}+\mathcal{L}_{FM}+\mathcal{L}_{P}$	28.142	0.802
$\mathcal{L}_{GAN}+\mathcal{L}_{FM}+\mathcal{L}_{1}$	27.392	0.834

注：加粗字体为每列最优的结果。





(a)未采用级联输出　　(b)采用级联输出

图 12　网络训练 20 epoch 时梯度分布

Fig. 12　Gradient distribution when the network is trained for 20 epochs

4　结束语

本文设计了一种基于条件生成对抗网络方法来解决单张梯级图像表面高光去除问题。通过将 SCFDB 模块替代为传统 U-net 架构基本模块来增强网络对于图像深层信息的特征提取能力，并设计出一种多尺度梯度级联方法赋能网络多尺度鉴别能力并稳定了网络训练梯度，通过对梯级表面模拟高光照射建立了一个用于网络训练和测试的自定义数据集。实验结果表明，本文模型相比基于颜色空间和先验条件分析的传统算法和基于 pix2pixHD 网络的方法在梯级图像数据集和经典高光数据集高光处理上都具有更优异的表现，具有一定的实际应用前景。但也能看到在处理高强度金属反光区域时所有方法都难以有效去除和还原梯级图像纹理，下一步的工作是继续优化网络结构，解决高强度反光去除问题，同时更进一步优化图像细节还原能力，提高算法效率。

参考文献：

[1] Lin S, Li Y, Kang S B, *et al.* Diffuse-specular separation and depth recovery from image sequences [C]// European Conference on Computer Vision. Berlin: Springer, 2002: 210-224.

[2] Lin S, Shum H Y. Separation of diffuse and specular reflection in color images [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2001, 1: I-I.

[3] Ren S, He K, Girshick R, *et al.* Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015, 28: 91-99.

[4] Yi R, Tan P, Lin S. Leveraging multi-view image sets for unsupervised intrinsic image decomposition and highlight separation [C]// Proc of AAAI Conference on Artificial Intelligence. 2020, 34 (07): 12685-12692.

[5] Suárez P L, Sappa A D, Vintimilla B X, *et al.* Deep learning based single image dehazing [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 1169-1176.

[6] Dong Y, Liu Y, Zhang H, *et al.* FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing [C]// Proc of AAAI Conference on Artificial Intelligence. 2020, 34 (07): 10729-10736.

[7] Chen J, Chen J, Chao H, *et al.* Image blind denoising with generative adversarial network based noise modeling [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 3155-3164.

[8] Wang J, Li X, Yang J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 1788-1797.

[9] Liu D, Long C, Zhang H, *et al.* ARShadowGAN: Shadow generative adversarial network for augmented reality in single light scenes [C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 8139-8148.

[10] Karnewar A, Wang O. Msg-gan: Multi-scale gradients for generative adversarial networks [C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2020: 7799-7808.

[11] Tan R T, Ikeuchi K. Separating reflection components of textured surfaces using a single image [M]// Digitally Archiving Cultural Objects. Boston: Springer, 2008: 353-384.

[12] Yang Q, Tang J, Ahuja N. Efficient and robust specular highlight removal [J]. IEEE Trans on pattern analysis and machine intelligence, 2014, 37 (6): 1304-1311.

[13] Yang Q, Wang S, Ahuja N. Real-time specular highlight removal using bilateral filtering [C]// European Conference on Computer Vision. Berlin: Springer, 2010: 87-100.

[14] Ramos V S, Júnior L G D Q S, Silveira L F D Q. Single image highlight removal for real-time image processing pipelines [J]. IEEE Access, 2019, 8: 3240-3254.

[15] Takahisa Yamamoto and Atsushi Nakazawa. General improvement method of specular component separation using high-emphasis filter and similarity function [J]. ITE Trans on Media Technology and Applications, 7 (2): 92–102, 2019. 6, 7.

[16] Xin Y, Jia Z, Yang J, *et al.* Specular reflection image enhancement based on a dark channel prior [J]. IEEE Photonics Journal, 2021, 13 (1): 1-11.

[17] 郭碧茹, 孔韦韦, 陈斌. 导向滤波的高光去除改进算法 [J]. 计算机工程与应用, 2021, 57 (20): 229-235. (Guo Bilu, Kong Weiwei, Chen Bin. Improved specular removal algorithm for guided filtering [J]. Computer Engineering and Applications, 2021, 57 (20): 229-235.)

[18] Lin S, Shum H Y. Separation of diffuse and specular reflection in color images [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2001, 1: I-I.

[19] Shah M Z A, Marshall S, Murray P. Removal of specular reflections from image sequences using feature correspondences [J]. Machine Vision and Applications, 2017, 28 (3): 409-420.

[20] 温佩芝, 周迎, 苗渊渊, 冯丽园. 多视角图像特征匹配的高光去除方法 [J]. 计算机工程与应用, 2018, 54 (23): 156-161. (Weng Peizhi, Zhou Ying, Miao Yuanyuan. A highlight removal method for multi-view image feature matching [J]. Computer Engineering and Applications, 2018, 54 (23): 156-161.)

[21] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, *et al.* Generative Adversarial Networks [J]. Advances in Neural Information Processing Systems, 2014, 3: 2672-2680.

[22] Mirza M, Osindero S. Conditional generative adversarial nets [J/OL]. (2014-11-06) [2021-07-23]. https://arxiv. org/abs/1411. 1784

[23] Isola P, Zhu J Y, Zhou T, *et al.* Image-to-image translation with conditional adversarial networks [C]// Proc of IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Press, 2017: 1125-1134.

[24] Wang T C, Liu M Y, Zhu J Y, *et al.* High-resolution image synthesis and semantic manipulation with conditional gans [C]// Proc of the IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Press, 2018: 8798-8807.

[25] Fu G, Zhang Q, Lin Q, *et al.* Learning to detect specular highlights from real-world images [C]// Proc of the 28th ACM International Conference on Multimedia. New York: ACM Press, 2020: 1873-1881.

[26] Chen Q, Xu J, Koltun V. Fast image processing with fully-convolutional networks [C]// Proc of IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 2017: 2497-2506.

[27] Huang G, Liu Z, Van Der Maaten L, *et al.* Densely connected convolutional networks [C]// Proc of IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Press, 2017: 4700-4708.

[28] Suo J, An D, Ji X, *et al.* Fast and high quality highlight removal from a single image [J]. IEEE Trans on Image Processing, 2016, 25 (11): 5441-5454.

[29] Yi R, Tan P, Lin S. Leveraging multi-view image sets for unsupervised intrinsic image decomposition and highlight separation [C]// Proc of AAAI Conference on Artificial Intelligence. 2020, 34 (07): 12685-12692.

[30] Dong H, Pan J, Xiang L, *et al.* Multi-scale boosted dehazing network with dense feature fusion [C]// Proc of IEEE/CVF conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Press, 2020: 2157-2167.